



Estimating Satellite Attitude From Pushbroom Sensors

Régis Perrier, Élise Arnaud, Peter Sturm, Mathias Ortner

► To cite this version:

Régis Perrier, Élise Arnaud, Peter Sturm, Mathias Ortner. Estimating Satellite Attitude From Pushbroom Sensors. CVPR 2010 - 23th IEEE Conference on Computer Vision and Pattern Recognition, Jun 2010, San Francisco, United States. pp.591-598, 10.1109/CVPR.2010.5540160 . inria-00514483

HAL Id: inria-00514483

<https://inria.hal.science/inria-00514483>

Submitted on 2 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating satellite attitude from pushbroom sensors

Régis Perrier*
INRIA Grenoble

Elise Arnaud*
Université Joseph Fourier, LJK

Peter Sturm*
INRIA Grenoble

Mathias Ortner
EADS Astrium

* 655, avenue de l'Europe 38330 Montbonnot, France
{firstname.lastname}@inrialpes.fr

Abstract

Linear pushbroom cameras are widely used in passive remote sensing from space as they provide high resolution images. In earth observation applications, where several pushbroom sensors are mounted in a single focal plane, small dynamic disturbances of the satellite's orientation lead to noticeable geometrical distortions in the images. In this paper, we present a global method to estimate those disturbances, which are effectively vibrations. We exploit the geometry of the focal plane and the stationary nature of the disturbances to recover undistorted images. To do so, we embed the estimation process in a Bayesian framework. An autoregressive model is used as a prior on the vibrations. The problem can be seen as a global image registration task where multiple pushbroom images are registered to the same coordinate system, the registration parameters being the vibration coefficients. An alternating maximisation procedure is designed to obtain Maximum a Posteriori estimates (MAP) of the vibrations as well as of the autoregressive model coefficients. We illustrate the performance of our algorithm on various datasets of satellite imagery¹.

was especially true in the past when digital cameras were first appearing; still nowadays, cameras used for earth observation usually record up to 25, 000 pixels in a single line.

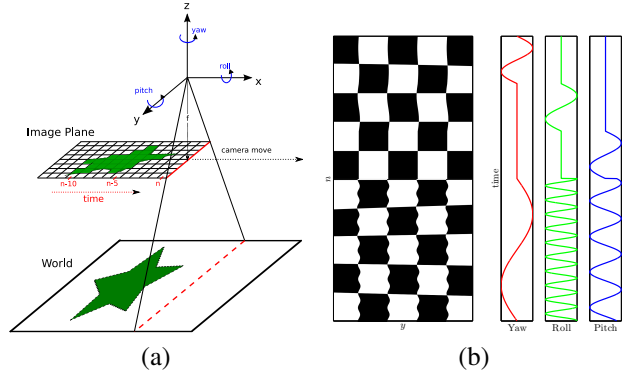


Figure 1. (a) Pushbroom acquisition principle: the camera is moving straight along the x axis and recording 1-D images over time denoted by n ; y is the camera axis and z the orthogonal axis to the image plane. We define the orientation of the camera with the yaw (rotation about z), the roll (rotation about x) and the pitch (rotation about y). (b) Example of warps in a regular checkerboard when the pushbroom camera is tilting around its 3 rotation axes.

1. Introduction

Over the last decade, pushbroom sensors have found increasing utility in several applications, ranging from passive remote sensing for earth observation, 3D inspection of containers for security, medical scanners for x-ray imagery to the personal flatbed scanner. Concerning its principle, the pushbroom camera is a linear sensor that takes 1-D images at several time instants. The sensor sweeps out a region of space; stitching together all 1-D images gives a complete 2-D image of the observed scene. This acquisition process is summarised in figure 1(a). As a result of this process, pushbroom sensors have the advantage of providing higher resolution images than classical perspective cameras. This

The recorded image is therefore a concatenation of 1-D images. This image is a non-distorted view of the observed scene only if the camera is moving at a constant speed and orientation (distortion in the pushbroom case refers to geometrical warps due to camera motion). Otherwise, geometric warps influence the acquisition process, and the image has to be corrected to properly represent the scene. This correction can be performed once the camera poses are estimated. When the sensor is fixed on a motorised support that ensures a constant speed and orientation, it is possible to estimate the position and orientation of the camera, assuming that its calibration parameters are constant over time [9, 6].

In airborne and spaceborne imagery, such assumptions of constant motion and orientation are violated since the craft is exposed to dynamic disturbances during the acquisition process. Those vibrations are caused by the engines as well as atmospheric and space turbulence. This becomes

¹This work was funded by EADS Astrium (European aerospace company and satellite manufacturer)

especially true with the recent manufacture of small satellites whose high sampling resolution makes these dynamic perturbations even more probable and critical. Even if a drift in the platform position can be recorded by GPS sensors and thus rectified, variation of the attitude² is harder to detect and leads to noticeable geometric distortions in the acquired image. Figure 1(b) shows an example of warps in the image when the camera orientation is changing over time; we define the three components of the attitude, the yaw, the roll and the pitch as the rotations around the x , the y and the z axis respectively, as shown in figure 1(a).

Current solutions use inertial sensors to retrieve the attitude of the platform [3, 16]. Apart from being very costly for space applications, the sampling frequency of such sensors (4Hz) is usually lower than the sampling frequency of each 1-D image of the pushbroom camera (2, 500Hz). Thus, high frequency disturbances cannot be recorded by those sensors and may not be rectified on the image.

In this paper, we present an original method to estimate the attitude variations of the pushbroom camera during the acquisition process in order to rectify each 1-D image. To do so, we rely on information acquired by the camera only. We exploit the specific geometry of the spaceborne platform where several pushbroom cameras are mounted in a same focal plane to get images of different modalities (section 3). The estimation process is embedded in a Bayesian framework, allowing us to combine an image data term and a prior vibration model (section 4). The data term describes a global registration of all images, the registration parameters being the vibration coefficients. The prior model accounts for the stationary properties of the vibrations and is expressed by an autoregressive model. The resulting algorithm (section 5) is an alternating maximisation procedure, whose all parameters are automatically estimated. This procedure is designed to obtain Maximum a Posteriori estimates (MAP) of the vibrations as well as of the autoregressive model coefficients. The knowledge of vibrations makes possible the rectification of the acquired images, while the knowledge of the autoregressive coefficients allows us to retrieve the vibrations in the case of unreliable image data such as clouds and sea.

2. Related work

Pushbroom cameras have been much less studied than the classical perspective cameras. Few works have been done on pushbroom camera calibration; [9] is probably one of the most important as the authors derive a simple model which allows to relate several pushbroom cameras observing a common scene with a matrix analogous to the fundamental matrix. More recently, a linear method that uses

homographies induced by the images of a planar object has proved to be simple and efficient for estimating calibration parameters [6]. These approaches cannot be applied to the context of spaceborne applications for two main reasons: first, these methods relate to the configuration of a sensor mounted on a fixed platform while the spaceborne platforms undergo vibrations; second, a large part of the calibration parameters such as the position of the satellite or the focal length are known, and the remaining parts to be estimated are the vibrations themselves.

Therefore, in the context of earth observation, most of the research focuses on fusing inertial sensors mounted on the craft to get a good estimate of the camera orientation [3, 16]. However, such sensors do not guarantee that vibrations are fully estimated, as their sampling frequency is lower than the frequency of acquisition of 1-D images. The possibility of using only image information to accurately retrieve the satellite vibrations is of great interest for satellite manufacturers as it would reduce the price of the craft.

To our best knowledge, there is only one paper that describes a solution to estimate the vibrations directly from images [4]. This approach is based on a local feature extraction on each 1-D image, that are further put into correspondence using correlation techniques for every pair of 1-D images. This allows the estimation of a relative vibration for each pair. The absolute vibration is finally recovered using a Wiener filtering process. As this algorithm is based on a local principle, it suffers from four main drawbacks: (a) this method does not account for the linear geometry of the camera as each feature in a same line of pixels is treated independently; (b) the case of no detected features can occur when the satellite is orbiting over textureless regions like sea, mountains with snow or clouds; (c) priors or sensor information cannot be easily added to the estimation process and (d) this procedure does not estimate directly the absolute attitude vibrations, as it first produces the set of relative vibrations that have to be filtered in a second step. Each relative vibration is calculated from each image pair and the global coherence is not ensured.

We believe that the problem of estimating absolute vibrations from image information can be handled with more appropriate methods, which are known to be effective in domains like super-resolution or optical flow estimation. First, we use pixel-based methods [13, 10, 1, 18] to register images, as they are more suited to estimate sub-pixel motion making optimal use of every pixel in the image [18]. Then the registration process is performed over all images in a global manner as this has been shown to give more accurate results for super-resolution applications [7]. Finally, as vibrations mainly originate from the engines of the spacecraft, we use an autoregressive prior [5, 14] to retrieve the warps between images.

²usual name for the orientation of the air and space vehicle in flight dynamics science defined by the yaw, the roll, and the pitch

3. Image acquisition system

In spaceborne applications, several pushbroom cameras are fixed in parallel on the focal plane of the satellite. Those cameras record different spectral bands, which are commonly red, blue, green and panchromatic bands for earth observation. Due to the geometry of the focal plane, the four cameras are not observing the same part of the scene at the same time instant, as can be seen in the figure 2. In this example, what is seen by the blue camera at time n is supposed to be seen at time $n + \tau_{12}$ by the panchromatic one. We can also notice lines of sight which are defined as projections of the linear sensor onto the observed scene. Attitude changes of the focal plane during the acquisition lead to varying position of each line of sight; thus images become irregularly sampled. We will see in the next section that we can globally register the set of images using the geometry of the focal plane.

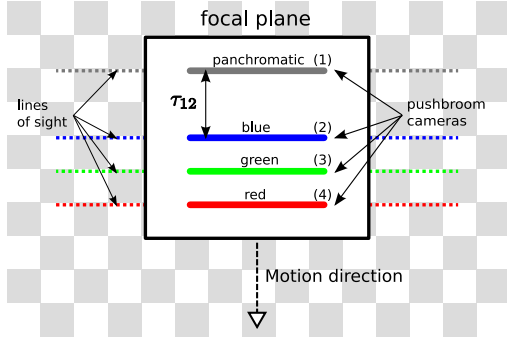


Figure 2. Standard focal plane geometry of observation satellite with 4 pushbroom cameras: panchromatic, blue, green and red (respectively enumerated as 1,2,3 and 4). What is seen by the camera 2 at time n will be seen by the camera 1 at time $n + \tau_{12}$.

This registration problem is rather difficult if only image information is used and no prior information. If we consider the case of a single pushbroom camera, then the problem is highly ill-posed: roll angles alone could be reasonably well estimated (in this special case, one can notice that the problem has similarities with the rolling shutter effect [12]), but variations in pitch and especially yaw, are difficult to recover. Methods that try to estimate all angles by maximising the similarity of neighboring pixels across successive 1-D images, typically result in making the dominant visible edges in the scene straight in the final rectified image, even if they correspond to curved roads. In our case we have two means for overcoming this: the presence of several pushbroom sensors makes all these angles observable by correlating all images. However, depending on the amount and especially the frequency of attitude variations, these correlations may or may not be sufficiently strong. Thus, we also use prior knowledge on the modeled phenomenon, here the assumption of its stationarity.

In this paper, we denote the set of 4 images (red, blue, green and panchromatic) as $\mathbf{I} = \{I_1, I_2, I_3, I_4\}$. Each image is defined by its pixel coordinates (y, n) in the pixel set \mathcal{S} , with $n \in [0, N - 1]$ being the discretised time and N the total number of time steps which is also equivalent to the number of lines in the acquired images. We call $\theta(n) \in \Theta$ the unknown attitude of the satellite at time n ; it is a (3×1) vector whose components are respectively the yaw $\theta_y(n)$, the roll $\theta_r(n)$ and the pitch $\theta_p(n)$. We call $\boldsymbol{\theta}$ the $(3N \times 1)$ vector that gathers all attitudes for all time instants. Also, as explained in the beginning, we use a model as a prior to infer the vibrations in a Bayesian setting; we denote by \mathbf{a} the vector of parameters for this model.

4. MAP formulation

The Bayesian formulation allows us to describe the problem in a suitable manner; we seek to maximise the *a posteriori* probability of the vibrations $\boldsymbol{\theta}$ and parameters \mathbf{a} given the observation of images \mathbf{I} . This yields:

$$\begin{aligned} (\hat{\boldsymbol{\theta}}, \hat{\mathbf{a}}) &= \underset{\boldsymbol{\theta}, \mathbf{a}}{\operatorname{argmax}} p(\boldsymbol{\theta}, \mathbf{a} | \mathbf{I}) \\ &= \underset{\boldsymbol{\theta}, \mathbf{a}}{\operatorname{argmax}} p(\mathbf{I}, \boldsymbol{\theta}, \mathbf{a}). \end{aligned}$$

The joint probability can also be expressed as:

$$p(\mathbf{I}, \boldsymbol{\theta}, \mathbf{a}) = p(\mathbf{I} | \boldsymbol{\theta}, \mathbf{a}) p(\boldsymbol{\theta} | \mathbf{a}) p(\mathbf{a}). \quad (1)$$

The prior over the autoregressive coefficients $p(\mathbf{a})$ is assumed to be uniform. We assume that the knowledge of vibrations fully explains the likelihood of images i.e. $p(\mathbf{I} | \boldsymbol{\theta}, \mathbf{a}) = p(\mathbf{I} | \boldsymbol{\theta})$. This term represents the likelihood of the images depending on the vibrations which is usually referred to as data term, whereas the second term in eq. (1) $p(\boldsymbol{\theta} | \mathbf{a})$ represents a prior on the vibrations.

4.1. Image data term

For the sake of clarity, we will first assume that all images are of the same modality (in the sense that images are radiometrically calibrated). Our strategy to handle the multi-modal case will be discussed in the results section (section 6). Also, we start by comparing two images and will derive a global formulation thereafter.

Given I_1 and I_2 spaced by τ_{12} in the focal plane (see figure 2), if the two images are perfectly aligned, i.e. if the satellite does not undergo any vibrations, we expect to have the following relationship:

$$\left[I_1(y, n) - I_2(y, n + \tau_{12}) \right] \sim \mathcal{N}(0, \sigma_i^2), \quad (2)$$

where σ_i^2 is the variance of a zero mean Gaussian acquisition noise i.i.d. over all pixels in images. As explained

earlier, a small variation in the satellite attitude induces a geometric distortion in the images. Formally, this distortion can be expressed by the use of a warp function W that maps the pixel coordinates to a new position depending on the attitude of the pushbroom camera [1]:

$$W : \mathcal{S} \times \Theta \rightarrow \mathcal{S}$$

$$W(y, n; \theta(n)) = (y', n').$$

Since vibrations occur at every time instant, eq. (2) has to be modified to better reflect the observed images:

$$\left[I_1(W(y, n; \theta(n))) - I_2(W(y, n + \tau_{12}; \theta(n + \tau_{12}))) \right] \sim \mathcal{N}(0, \sigma_i^2).$$

Fixing I_1 as the reference image, this expression is equivalent to the following one:

$$\left[I_1(y, n) - I_2(W(y, n + \tau_{12}; \theta(n + \tau_{12}) - \theta(n))) \right] \sim \mathcal{N}(0, \sigma_i^2). \quad (3)$$

From the previous equation, one can notice that:

$$\begin{pmatrix} \theta_y(n + \tau_{12}) - \theta_y(n) = (\theta_y * k_{12})(n) \\ \theta_r(n + \tau_{12}) - \theta_r(n) = (\theta_r * k_{12})(n) \\ \theta_p(n + \tau_{12}) - \theta_p(n) = (\theta_p * k_{12})(n) \end{pmatrix},$$

where $*$ denotes the convolution operator and

$$k_{12} = \begin{bmatrix} -1 & \underbrace{0 \dots 0}_{\tau_{12}-1 \text{ zeros}} & 1 \end{bmatrix} \quad (4)$$

is a kernel built upon the time gap between the acquisition of I_1 and I_2 ; this time gap is also corresponding to the spatial gap between the sensors on the focal plane. This basically shows that trying to estimate θ from any pair of images $\{I_i, I_j\}$ acquired by the focal plane leads to the non blind deconvolution of θ by a kernel k_{ij} . This kernel depends on the time delay between the recording of I_i and I_j like in expression (4).

In a global formulation, let $I_i^{\tau_{ij}}(y, n) = I_i(y, n + \tau_{ij})$ be the shifted image in time with factor τ_{ij} , $\mathbf{y} = [y, n]$ the vector of pixel coordinates, and K_{ij} the matrix convolution operator with kernel k_{ij} , then the log-likelihood arises immediately:

$$\log(p(\mathbf{I}|\theta)) \propto \sum_{i,j; i \neq j} \sum_{\mathbf{y} \in \mathcal{S}} \left(I_i(\mathbf{y}) - I_j^{\tau_{ij}}(W(\mathbf{y}; K_{ij}\theta)) \right)^2 + \text{cst}. \quad (5)$$

This last equation is the expression of a pixel-based registration method, also known as the Lucas-Kanade method [13, 1]. The outer sum takes into account all possible pairs of images while the inner one is a summation over all pixels in the images. The additional constant term refers to the normalizing term of the Gaussian p.d.f. which is independent of θ , and thus of no importance for the minimization procedure. Let us stress that this expression relates the *absolute* vibrations to all image data, thus ensuring the global coherence of the estimation with respect to all image pairs.

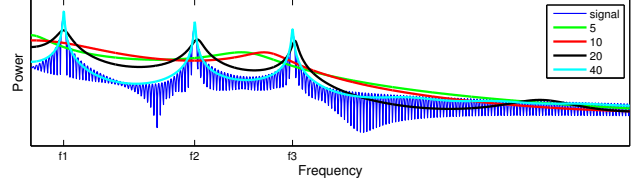


Figure 3. Autoregressive modeling of a signal (sum of three frequencies f_1 , f_2 and f_3) in the frequency domain for several orders ($p = 5, 10, 20, 40$).

4.2. Vibration prior model

Variations in the attitude of the satellite are mainly due to its engines. Modelling such a vibration process is a challenging problem in the signal processing literature and more particularly in monitoring of structural and mechanical systems [5, 2]. Typical approaches use time series, time-frequency and modal analysis to detect a failure in a mechanical system where several engines are rotating and thus vibrating; the main assumption on the vibration process is its stationarity, in the sense that it keeps the values of its harmonic frequencies constant over time. Thus, a failure is recorded if those stationary properties are changed.

One effective and commonly used solution is the autoregressive model where a signal (vibration) is modeled as a linear combination of its past values [14, 15]. This yields for $n \in [0, N - 1]$:

$$\theta(n) = - \sum_{k=1}^p a_k \theta(n - k) + \epsilon(n), \quad (6)$$

where a_k for $1 \leq k \leq p$, are the parameters of the model and p its order. a_k is a (3×1) vector whose components account for the yaw, the roll and the pitch, and \mathbf{a} is the $(3p \times 1)$ vector gathering all a_k : $\mathbf{a} = [a(1) : a(p)]^T$. $\epsilon(n)$ is taken here to be a zero mean i.i.d. Gaussian noise with variance σ_a^2 . This model is linked to linear filtering theory and is widely used to predict or filter stationary time series [14, 19, 11]. In the frequency domain, it is easier to see the effect of the autoregressive model which tries to fit the spectrum of the signal depending on its order as can be seen on figure 3.

Let A be the $(N - p \times N)$ matrix operator with the autoregressive coefficients:

$$A = \begin{bmatrix} a(p) & a(p-1) & \dots & a(1) & -1 & 0 & \dots & 0 \\ \vdots & & & \ddots & & & & \vdots \\ 0 & \dots & 0 & a(p) & a(p-1) & \dots & a(1) & -1 \end{bmatrix}$$

Thus the logarithm of the *a priori* vibration model can be expressed as:

$$\log(p(\theta|\mathbf{a})) \propto \|A\theta\|^2 + \text{cst}, \quad (7)$$

where the constant term accounts for the normalizing factor of the Gaussian p.d.f..

5. Algorithm

Our goal is to estimate realisations of θ and \mathbf{a} that maximise the defined *a posteriori* distribution $p(\theta, \mathbf{a}|\mathbf{I})$. However, this global optimization problem has in general no straightforward solution. Thus, we consider instead an iterative approach consisting in maximising the posterior probability alternatively in the first and second variable. Starting from current estimates $\theta^{(t)}$ and $\mathbf{a}^{(t)}$ at iteration t , we consider the following two-step updating:

I. Estimation of $\mathbf{a}^{(t+1)}$

$$\mathbf{a}^{(t+1)} = \underset{\mathbf{a}}{\operatorname{argmax}} p(\theta^{(t)}, \mathbf{a}|\mathbf{I}) = \underset{\mathbf{a}}{\operatorname{argmax}} p(\theta^{(t)}|\mathbf{a}).$$

II. Estimation of $\theta^{(t+1)}$

$$\begin{aligned} \theta^{(t+1)} &= \underset{\theta}{\operatorname{argmax}} p(\theta, \mathbf{a}^{(t+1)}|\mathbf{I}) \\ &= \underset{\theta}{\operatorname{argmax}} p(\mathbf{I}|\theta) p(\theta|\mathbf{a}^{(t+1)}). \end{aligned}$$

Convergence to a local maximum is guaranteed since the posterior probability is increasing at each step:

$$p(\theta^{(t)}, \mathbf{a}^{(t)}|\mathbf{I}) \leq p(\theta^{(t)}, \mathbf{a}^{(t+1)}|\mathbf{I}) \leq p(\theta^{(t+1)}, \mathbf{a}^{(t+1)}|\mathbf{I}). \quad (8)$$

The first step consists in estimating the autoregressive coefficients $\mathbf{a}^{(t)}$ from the current vibration estimate. This is done using the autocorrelation method, also known as the Yule-Walker equation [14]. The order of the model p is chosen using the Akaike criterion [14].

Deduced from eq. (5) and (7), the second step consists in the minimization over θ of the following equation:

$$\sum_{i,j;i \neq j} \|I_i(\mathbf{y}) - I_j^{\tau_{ij}}(W(\mathbf{y}; K_{ij}\theta))\|^2 + \lambda \|A\theta\|^2, \quad (9)$$

where λ is a real scalar proportional to the fraction between the noise variance in the data and the noise variance of the model; this can also be seen as a trade-off between the data and the model term. In practice, a cross validation method is used to estimate this parameter. Eq. (9) has to be minimised iteratively due to its non linear form. We choose a Gauss-Newton algorithm to estimate θ for its good performances in registration methods [1]. However, any other non linear least squares method could be used. Using a gradient descent procedure does not necessarily require the knowledge of an analytical form of the warp W as long as derivatives of the warp $\frac{\partial W}{\partial y}$ and $\frac{\partial W}{\partial n}$ are available. In the pushbroom case, we can derive an analytical expression of W (which is quite similar to the perspective case) as long as the observed scene is flat. In the opposite case when the scene has ground elevation, the warp function depends on the scene and the expression of W is much more complex. In our experiments, we use numerical derivatives of W ; this has several advantages as Digital Elevation Model (DEM) and

Initialisation:

- $\mathbf{a}^{(0)} = \mathbf{0}, \lambda^{(0)} = 0, p^{(0)} = 0, \mathbf{I}^{(0)} = \mathbf{I}$
- $\Delta\theta^{(0)} = \text{minimize_1step_Eq9}(\mathbf{I}^{(0)}, \mathbf{a}^{(0)}, \lambda^{(0)})$
- $\theta^{(0)} = \Delta\theta^{(0)}$

Set $t=1$

- I:
1. $p^{(t)} = \text{select_order}(\theta^{(t-1)})$
(Akaike criterion)
 2. $\mathbf{a}^{(t)} = \text{select_AR_coefficients}(\theta^{(t-1)}, p^{(t)})$
(autocorrelation method)
 3. if $t = 1$, then $\lambda^{(t)} = \text{select_parameter}(\mathbf{I}, \mathbf{a}^{(t)})$
(cross validation)
else $\lambda^{(t)} = \lambda^{(t-1)}$
- II:
4. $\Delta\theta^{(t)} = \text{minimize_1step_Eq9}(\mathbf{I}^{(t-1)}, \mathbf{a}^{(t)}, \lambda^{(t)})$
 5. $\theta^{(t)} = \theta^{(t-1)} + \Delta\theta^{(t)}$
 6. $\mathbf{I}^{(t)} = \text{warp_images}(\mathbf{I}, \theta^{(t)})$
 7. $t = t + 1$, till $\Delta\theta^{(t)} < \epsilon$

Figure 4. Algorithm

optical distortions of the lens can be directly added to those numerical derivatives. Finally, the whole algorithm can be summarized in the scheme on figure 4.

One can notice that in practice, the algorithm computes only one step of the gradient descent at step 4 of the algorithm. Nevertheless, the algorithm convergence is ensured as equation (8) still holds.

6. Experimental results

In this section, experimental results are presented to highlight the relevance of our algorithm. Two satellite datasets are studied. They both have been simulated by EADS Astrium, meaning that the ground truth on vibrations is available. The simulation process creates real-life condition data as it takes into account effects from camera acquisition noise, ground elevation, radiometric distortions, and mechanical vibration conditions. This data can therefore be considered as difficult as could be real data.

Each dataset is composed of 4 multispectral images: green, red, blue and panchromatic. All images are of size (2564×900) pixels; the pushbroom sensor size being 900 pixels. For all experiments we use a Matlab implementation on a Core2 duo at 3GHz with 3.8GiB. The algorithm converges in 4 iterations and we obtain a computational time below 100 seconds for both datasets.

To handle the different modalities of images, we tried two approaches; the first is to prefilter images with a high-pass filter to extract dominant information, and the second is to correct radiometric differences between images using

a linear model before the registration process. Both showed better results than processing raw images directly. However, the second method performed slightly better than the first one. We presume two reasons according to that observation. High-pass filtering usually increases the noise level and as a consequence can decrease the accuracy of the registration process. Moreover, panchromatic and RGB modalities are quite well correlated by default contrary to, for example, infrared or radar modalities. Thus, a linear radiometric model is enough in this case to correct images. We use the second solution to present the following results.

The first dataset corresponds to a focal plane where cameras are equally spaced by 20 time samples, which is equivalent to 20 lines of pixels in the images. During this experiment, λ is estimated to be 9.10^{11} and the autoregressive order p is in a range between [294, 320] depending on the iteration.

Figure 5 shows image registration results, the roll and pitch estimations compared to the ground truth over time, and the autoregressive prior spectrum compared to the ground truth spectrum at the first and last iterations. As can be seen on figure 5, the error on the vibration estimation is negligible: the standard deviation between the estimate and the ground truth is below $\frac{7}{100}$ in pixel unit, either for the roll and the pitch. One can notice on figure 5 that from the first iteration, the shape of the prior is fitting well the ground truth spectrum, though one low frequency is missing. This low frequency is recovered in the following iterations as can be seen in the final autoregressive spectrum.

Let us remark that no estimation of the yaw is given, as in that experiment, the yaw causes pixel deformations lower than $\frac{2}{100}$ pixel on each vertical side of the images (litterature shows at best pixel accuracy of $\frac{1}{100}$ in mono-modal and $\frac{5}{100}$ in multi-modal cases [8, 17]). These deformations are dominated by geometrical distortions induced by the roll and pitch. The same effect can be observed for the second experiment.

The second example shows a challenging case as the vibration signal has more low frequency components. Due to the convolution kernels k_{ij} (Eq. (4)), most of the low frequency information is lost and the deconvolution process becomes highly ill posed. In this dataset, cameras are spaced by 33.5, 40 and 20 time samples following the scheme in figure 2. In that case, λ is estimated to be 10^{10} and the autoregressive order p is in a range between [169, 205] depending on the iteration. Figure 6 shows image registration results and attitude estimates as in the previous dataset, as well as the spectrum of the autoregressive model for the first and last iterations. The standard deviation between the estimate and the ground truth is below $\frac{6}{100}$ in pixel unit, either for the roll and the pitch. Our algorithm still performs well as can be seen in the residual curves (drawn in red in figure 6). Thus, the autoregressive prior helps reg-

ularizing the estimate in those frequencies. Still, we can see that the autoregressive model gives a rough estimate of the low frequencies whereas higher frequencies estimate is much more accurate.

These experimental results demonstrate the performance of our algorithm. The autoregressive prior regularizes the solution even though the low frequencies are lost in the convolution process, and is particularly accurate in the high frequencies. In general we notice that the method presents very good results in high frequency estimation, which can not be obtained by the use of inertial sensors alone, as they only provide low frequency information. We still need to conduct more experiments on lower frequencies with broader pixel warps. Being more accurate on low frequencies using the autoregressive prior would imply to work on much longer time acquisition of images. Learning those coefficients to get a better prior is a part of our future work. Also, our next step will be to fuse measurements from images and inertial sensors so as to get the best estimate of attitude variations for all frequencies. The Bayesian setting we used easily allows us to combine prior as well as observations by other sensors in a coherent manner.

7. Conclusion

In this paper, we have presented a novel and efficient algorithm for estimating attitude variations of pushbroom cameras, that can be applied in spaceborne imagery. The definition of such an algorithm, that demonstrates the possibility of using image information to accurately retrieve the satellite vibrations, is of great interest for satellite manufacturers. The very promising results have demonstrated the interest of combining a global registration technique on all images to an autoregressive model for the estimation of the absolute vibrations. The estimation of the prior coefficients is one of the algorithm's key-points as it enables the image correction in case of difficult data like clouds, or sea. In a future work, the flexibility of our Bayesian setting will enable us to investigate the fusion of information from images and inertial sensors.

References

- [1] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(1):221 – 255, 2004. 2, 4, 5
- [2] E. P. Carden and P. Fanning. Vibration Based Condition Monitoring: A Review. *Structural Health Monitoring*, 3(4):355–377, 2004. 4
- [3] M. Cramer, D. Stallmann, and N. Haala. Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications. In *International Archives of Photogrammetry and Remote Sensing, Vol. 33 Part B3*, pages 198–205, 2000. 2

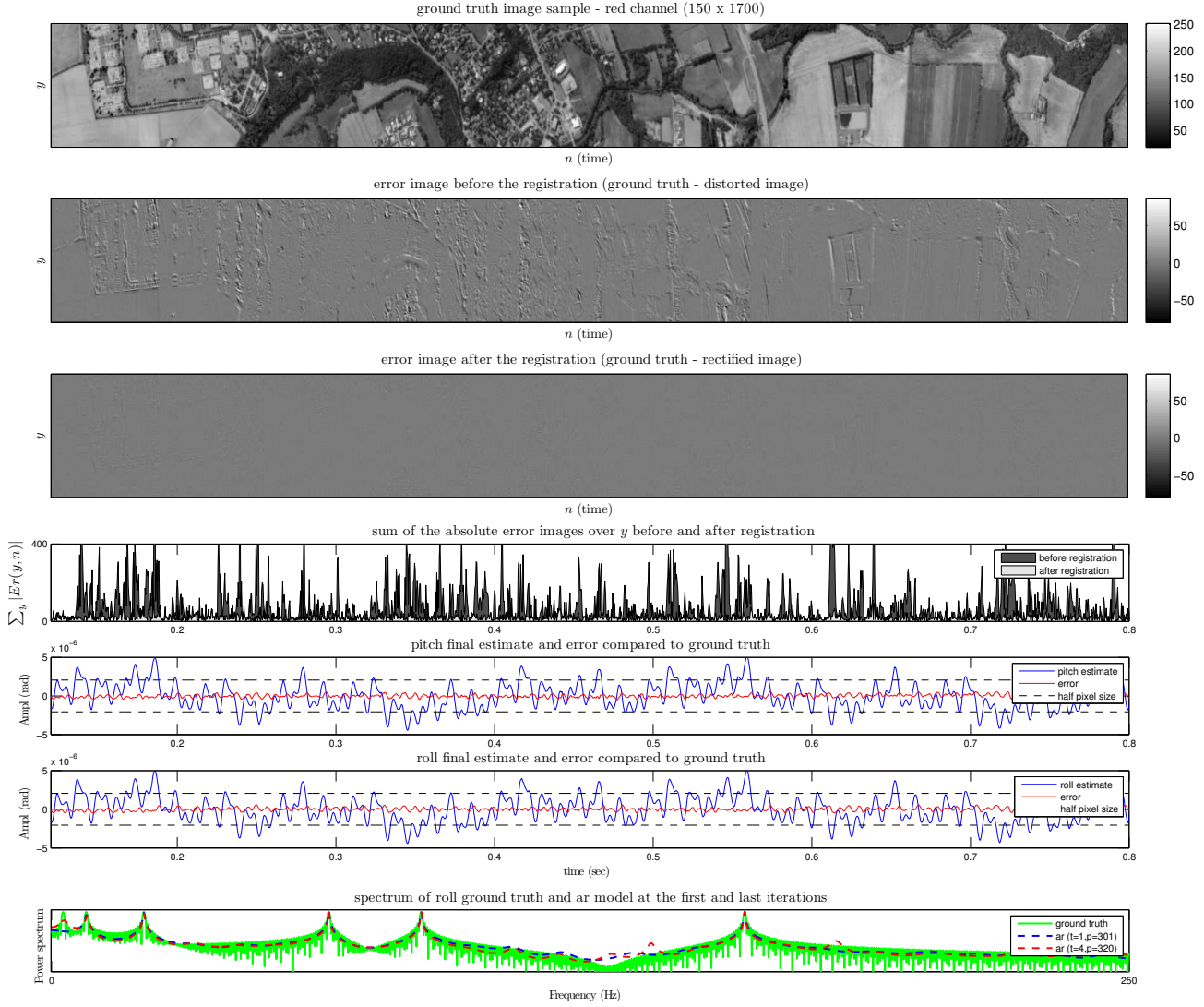


Figure 5. Dataset 1 results: the top figure is a (150×1700) image patch of the panchromatic camera. Below, the following two figures show intensity differences between the real image and the acquired image before and after registration. Then, the first graph presents the sum over y of absolute error image before and after registration depending on the time. The next two graphs draw attitude variation estimate in radian for the roll and the pitch in blue and the error compared to ground truth in red. The last figure is the spectrum of the autoregressive model at the first and last iterations compared to the ground truth spectrum in green.

- [4] F. de Lussy, D. Greslou, and L. Gross Colzy. Process line for geometrical image correction of disruptive microvibrations. In *International Society for Photogrammetry and Remote Sensing*, pages 27–35, 2008. 2
- [5] S. Doebling, C. Farrar, M. Prime, and D. Shevitz. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review. Technical report, Los Alamos National Lab., 1996. 2, 4
- [6] J. Drareni, P. Sturm, and S. Roy. Plane-Based Calibration for Linear Cameras. In *The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras - OMNIVIS*, 2008. 1, 2
- [7] S. Farsiu, M. Elad, and P. Milanfar. Constrained, globally optimal, multi-frame motion estimation. In *Proc. of the 2005 IEEE Workshop on Statistical Signal Processing*, pages 1396 – 1401, 2005. 2
- [8] W. Förstner. Quality assessment of object location and point transfer using digital image correlation techniques. In *International Archives of Photogrammetry*, 1984. 6
- [9] R. Gupta and R. I. Hartley. Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):963–975, 1997. 1, 2
- [10] M. Irani and P. Anandan. About direct methods. In *Proceedings of the International Workshop on Vision Algorithms*, pages 267–277, 1999. 2

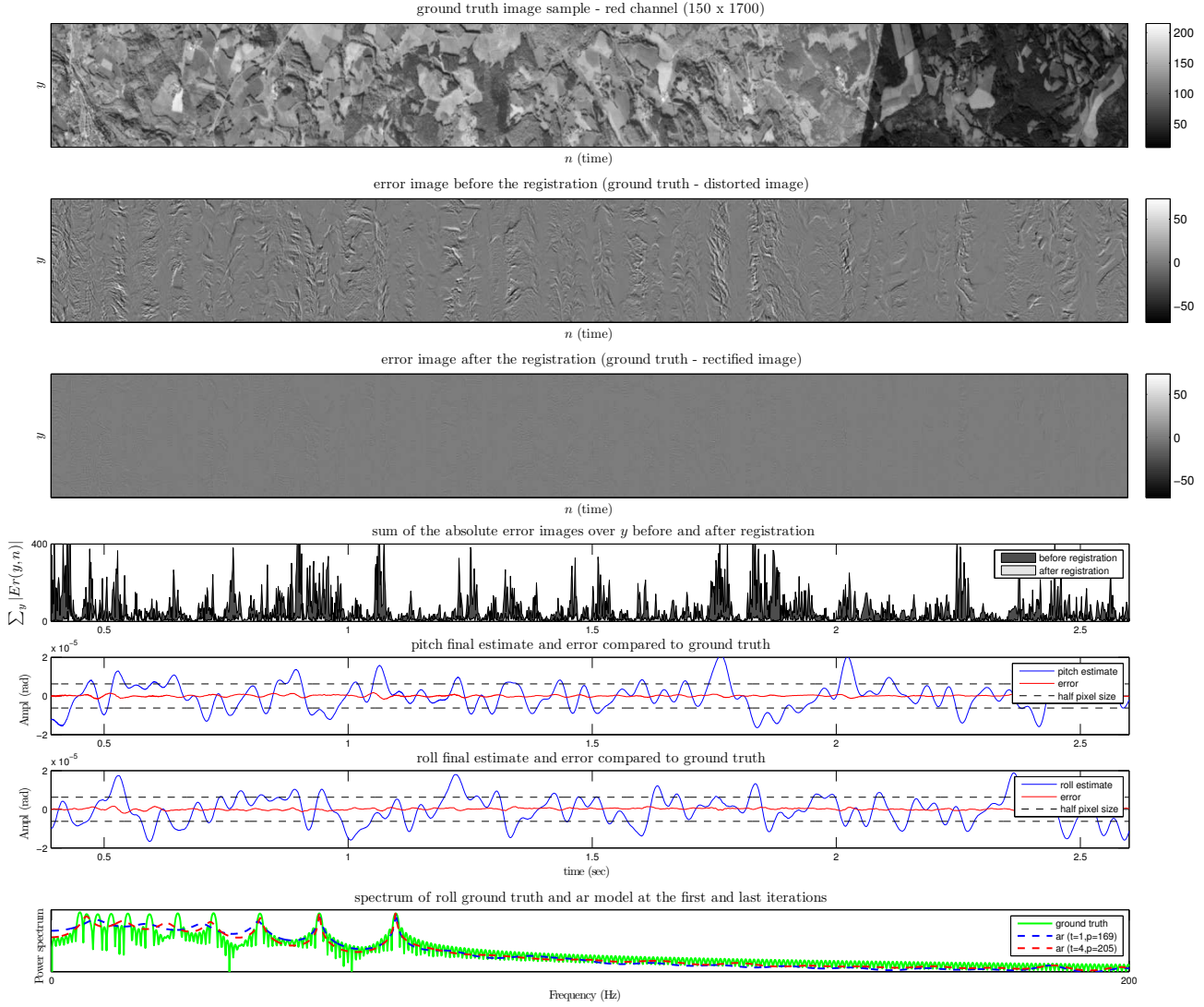


Figure 6. Dataset 2 results: the top figure is a (150×1700) image patch of the red camera. Below, the following two figures show intensity differences between the real image and the acquired image before and after registration. Then, the first graph presents the sum over y of absolute error image before and after registration depending on the time. The next two graphs draw attitude variation estimate in radian for the roll and the pitch in blue and the error compared to ground truth in red. The last figure is the spectrum of the autoregressive model at the first and last iterations compared to the ground truth spectrum in green.

- [11] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960. 4
- [12] C.-K. Liang, Y.-C. Peng, and H. Chen. Rolling shutter distortion correction. In *Visual Communications and Image Processing 2005*, volume 5960. SPIE, 2005. 3
- [13] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pages 674–679, 1981. 2, 4
- [14] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975. 2, 4, 5
- [15] S. L. Marple. *Digital spectral analysis: with applications*. Prentice-Hall, Inc., 1986. 4
- [16] D. Poli. General model for airborne and spaceborne linear array sensors. In *International Archives of Photogrammetry and Remote Sensing*, volume 34, 2002. 2
- [17] D. Robinson and P. Milanfar. Fundamental performance limits in image registration. *IEEE Transactions on Image Processing*, 13:1185–1199, 2003. 6
- [18] R. Szeliski. Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104, 2006. 2
- [19] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964. 4